ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет Санкт-Петербургская школа физико-математических и компьютерных наук Национального исследовательского университета

«Высшая школа экономики»

**Применение методов машинного обучения для прогнозирования геохимических характеристик озёр Арктического региона**

Курсовая работа

по направлению подготовки 01.04.02 Прикладная математика и информатика

образовательная программа «Анализ больших данных в бизнесе,

экономике и обществе»

Выполнила:

Санжапова Ляйсан Маратовна

студентка группы МБД-181

Научный руководитель:

преподаватель департамента прикладной

математики и бизнес-информатики

Белялов Ильдар Кафисович

**Оглавление**

[Введение 3](#_Toc11429769)

[1. Обзор литературы 5](#_Toc11429770)

[1.1. Факторный анализ в русскоязычных публикациях 5](#_Toc11429771)

[1.2. Факторный анализ в зарубежных публикациях 6](#_Toc11429772)

[1.3. Методы машинного обучения 7](#_Toc11429773)

[2. Описание и подготовка данных 9](#_Toc11429774)

[2.1. Описание имеющихся данных по 3 озёрам Мурманской области 9](#_Toc11429775)

[2.2. Описание имеющихся данных по озёрам Арктического региона 9](#_Toc11429776)

[3. Описание применяемых методов 12](#_Toc11429777)

[3.1. Разведывательный факторный анализ 12](#_Toc11429778)

[3.2. Метод опорных векторов 13](#_Toc11429779)

[4. Анализ данных и результаты 14](#_Toc11429780)

[4.1. Факторный анализ 3 озёр Мурманской области 14](#_Toc11429781)

[4.2. Прогнозирование геохимических характеристик озёр Арктического региона 19](#_Toc11429782)

[Заключение 21](#_Toc11429783)

[Список литературы 22](#_Toc11429784)

[Приложение 25](#_Toc11429785)

# Введение

В современном мире, где стремительно развиваются разные инструменты интеллектуального анализа данных, всё более актуальной становится кросс-функциональность и интеграция имеющихся знаний в смежные области и дисциплины науки. Давно стало ясно, что уместное применение методов машинного обучения и других аспектов науки о данных способно значительно обогатить наши представления как об окружающем мире, так и о нас самих. Этот факт подтверждается многочисленными исследованиями, направленными на практическое применение инструментов анализа для изучения природных, психологических, социальных, политических и других процессов. Усовершенствование технических возможностей вычислительных машин позволило также ускорить и повысить качество проводимых манипуляций с данными, что делает анализ более доступным и качественным.

В своей работе я хочу сфокусировать внимание на применении методов машинного обучения для исследования природных процессов, а именно – для изучения взаимосвязи между гидро- и геохимическими показателями (Водяницкий, 2008) в водоёмах Арктического региона.

**Актуальность** данной работы заключается в том, что за последние годы было проведено множество исследований с применением статистических, математических и методов машинного анализа для решения различных задач геохимии, как в иностранных, так и российских научных кругах, однако среди них не было попыток предсказания геохимических характеристик по гидрохимическим показателям. Ценность подобных предсказаний заключается в том, что при достаточно высоком уровне качества они смогут отчасти упразднить необходимость проведения длительных дорогостоящих лабораторных анализов химического состава донных отложений, и отчасти подобное прогнозирование сможет быть использовано для предиктивного анализа и принятия превентивных мер по устранению, допустим, опасных для экосистемы концентраций веществ.

**Целью** данной работы является исследование возможности практического применения методов машинного обучения для прогнозирования геохимических характеристик водных объектов.

Для достижения поставленной цели сформулированы следующие **задачи**:

* Обеспечить взаимодействие с экспертами в области геохимии, получить от них вводные знания, постановку практической задачи и набор данных;
* Провести подробный литературный обзор существующих исследований на данную тему;
* Провести анализ имеющихся данных с применением одного из существующих подходов, а именно – факторный анализ;
* По возможности провести анализ имеющихся данных с применением более сложных методов машинного обучения;
* Проанализировать полученные результаты, сделать выводы.

**Объектом** в данном исследовании являются геохимические характеристики озёр Арктического региона, расположенных на Российской территории, а именно в Мурманской области.

Работа состоит из четырёх частей. В первой части приводится литературный обзор статей и публикаций на заданную тему, анализ применённых в них подходов и наборов данных. Во второй части приводится описание данных, в третьей части – описание применяемых методов. В последнем разделе приводится факторный анализ данных и прогнозирование геохимических характеристик методами машинного обучения, а также интерпретация полученных результатов.

# Обзор литературы

Для удобства разделим обзор на публикации в российских изданиях и зарубежных. По просьбе экспертов в области геохимии, с которыми было установлено взаимодействие, и от которых была получена постановка практической задачи, при обзоре исследований особенно сфокусируемся на тех, которые были проведены с применением факторного анализа. Также проведём обзор и тех, что применяли различные методы машинного обучения для решения задач геохимии.

## Факторный анализ в русскоязычных публикациях

В отечественной геохимии активно используются математические и статистические методы анализа данных, проводятся конференции и научные школы. По окончании подобных мероприятий выходят сборники, которые и являются основным источником статей для приведённого в данном разделе обзора.

Всероссийская (с международным участием) научная школа «Математические исследования в естественных науках» (сайт) проводится ежегодно, к нынешнему моменту имеется уже 15 сборников статей. Среди публикаций на стыке гидро- и геохимии за последние годы особенно выделяются исследования с применением факторного анализа донных отложений.

В публикации (Вокуева С. И., 2017) авторы выявляют наиболее значимые гидрохимические факторы, определяющие реакции на антропогенное загрязнение озера Имандра в Мурманской области. Для химического анализа использовались образцы диатомовых комплексов и донных отложений. В исследовании использован корреляционный, кластерный анализ и многомерное статистическое исследование методом главных компонент (principal component analysis, PCA). В результате РСА было выявлено и интерпретировано 2 основных фактора, на основании которых исследованные станции забора материалов были классифицированы согласно их расположению в двумерном факторном пространстве. Полученная классификация согласуется с результатами кластерного анализа.

В публикации (Мицуков А. С., 2018) также проводится исследование озера Имандра. Для химического анализа использовались только образцы донных отложений. В исследовании использовался корреляционный и факторный анализ. В результате было выявлено 2 техногенных фактора, влияющих на формирование химического состава донных отложений.

## Факторный анализ в зарубежных публикациях

В статье (Clemens Reimann, 2002) исследуют Европейскую Арктику, а именно материнскую породу подзолистых почв. Для химического анализа использовалось 605 образцов, из которых было получено 89 переменных по 50 химическим элементам. Для факторного анализа использовалась программа S-PLUS, в которой реализовался как факторный анализ (factor analysis, FA), так и анализ методом главных компонент РСА. Исследование проводится на исходных, логарифмированных и трансформированных по Бокс-Коксу данных.

Данное исследование ценно своим практическим применением и доступностью изложенных в статье особенностей геохимических данных, полученных выводов. Так, стоит отметить, что геохимические данные очень редко подчиняются нормальному распределению; есть major элементы, концентрация которых измеряется в десятках процентов (), minor элементы () и trace элементы (. Сумма всех параметров для одного образца сходится к 100%, что влияет на матрицу корреляций и конечную корректность результатов. В статье рекомендуется сначала логарифмировать данные, а затем стандартизировать их при необходимости. Также возможно трансформирование Бокса-Кокса, которое даёт результаты ближе к нормальному.

В статье (Shan Hu, 2013) проводится исследование концентрации фтора в составе подземных вод провинции Шэньси, Китай. Существует ряд возможных причин, обуславливающих высокое содержание фтора: пример из естественных – длительное взаимодействие воды с минералами и вымывание фтора из отвердевших масс, пример из искусственных – загрязнение применяемыми в агрохозяйстве удобрениями. Концентрация фтора зависит также от уровня кислотности или щёлочности воды, концентрации в ней таких элементов, как натрий и кальций, а также гидрокарбонатов. Для определения параметров, на практике влияющих на содержание фтора в воде, используют метод РСА. Данные представляют собой 116 наблюдений по 16 переменным.

В аналогичной статье (Yuxiao Jiang, 2015) проводится исследование концентрации мышьяка в составе подземных вод региона Хетао, Китай. Для определения параметров, на практике влияющих на содержание мышьяка в воде, используют метод РСА и иерархический кластерный анализ. Данные представляют собой 90 наблюдений по 22 переменным. Для стандартизации использовали Z-оценку.

В работе (Drouin, 2012) проводится исследование погребённых кимберлитов через анализ подземных вод. Для анализа используется прикладное программное обеспечение SPSS Statistics 18, в котором реализован метод РСА. Данные представляют собой набор из 53 и 37 наблюдений, которые отличаются глубиной сбора образцов, по 32 переменным. Примечательно, что после исключения выбросов оставшиеся наблюдения и переменные относились как 2:1, хотя рекомендовано отношение от 5:1, однако критерий адекватности выборки Кайзера-Мейера-Олкина и тест Бартлетта указали на статистическую значимость корреляции между переменными. В результате факторного анализа было выявлено 3 ярко выраженных фактора.

## Методы машинного обучения

В статье (Shoutao Jiao, 2018) приводится исследование определения возможности добычи горных ископаемых методами машинного обучения. Приводится широкий список актуальных статей и публикаций, очерчены основные этапы истории развития этой кросс-функциональной сферы.

В статье (Hazi Md. Azamathulla, 2010) исследуется осадочная нагрузка или концентрация взвеси в 3 реках Малайзии. Для предиктивного моделирования используется метод опорных векторов (support vector machine, SVM). В результате полученная модель характеризуется высоким качеством предсказаний, показывая коэффициент детерминации равный 0.958 и среднеквадратическое отклонение равное 0.0698, что значительно превосходит аналогичные показатели для традиционных моделей.

В русскоязычной статье (Буевич А. Г., 2016) приводится сравнение традиционных геостатистических методов с искусственными нейронными сетями (ИНС) для прогнозирования и визуализации пространственного распределения химических элементов в поверхностном слое почвы. Результаты исследования показали превосходство ИНС над традиционным методом.

В статье (Kunwar P.Singh, 2009) также применяются ИНС, однако для иной задачи – для расчёта уровней растворённого кислорода и биохимической потребности в кислороде в реке Гомти, Индия.

В статье (Musavi-Jahromi S.H., 2008) применяются ИНС для моделирования качества воды в реке Карун, Иран. Данные представляют собой временной ряд с 1985 по 2006 год. Для моделирования используется нейронная сеть Qnet 2000.

# Описание и подготовка данных

Основные данные были получены благодаря сотрудничеству с заинтересованными представителями геохимической науки, в частности аспирантки СПБГУ и её научного руководителя.

Было получено 2 набора данных. Первый характеризует 3 озера Мурманской области и был получен благодаря непосредственным экспедициям, полевым сборам и химическим анализам, проведённым нашими коллегами. Второй набор данных по озёрам Мурманской области был получен из многочисленных каталогов.

## Описание имеющихся данных по 3 озёрам Мурманской области

Набор данных представляет из себя 3 таблицы: по первому озеру 12 записей, по второму – 29, по третьему – 13. Каждая запись отражает информацию по 47 параметрам: depth – глубина забора образца (см); Li, Be,, P, Sc, Ti, V, Cr, Mn, Co, Ni, Cu, Zn, Rb, Sr, Y, Zr, Nb, Mo, Ag, Cd, Sn, Sb, Cs, Ba, La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Hf, Ta, W, Tl, Pb, Bi, Th, U – концентрация соответствующего вещества в образце. Совмещённый датасет состоит, соответственно, из 54 записей по 47 параметрам. В Приложении А приведена описательная статистика совмещённого датасета. В Приложении Б представлена матрица корреляций совмещённого датасета.

## Описание имеющихся данных по озёрам Арктического региона

Второй набор данных по озёрам Мурманской области был получен из многочисленных каталогов, разбитых на районы: Юго-Восточный (Н.А. Кашулин, 2012), Восточный (Н.А. Кашулин, 2010), Центральный и Юго-Западный (Н.А. Кашулин, 2013).

Набор данных разделён на 2 части: датасет для обучения, который состоит из 92 записей по 30 параметрам, и датасет для предсказаний, который состоит из 428 записей по 18 параметрам. Первые 18 параметров в обоих датасетах одинаковы: pH – уровень кислотности, electricity – электропроводность (мкS/см), mineralization – общая минерализация (мг/л), alkalinity – щёлочность (мк-экв/л), colour – цветность (град.); Ca, Mg, Na, K, HCO3, SO4, Cl, NH4, NO3, N, PO4, P, Fe – концентрация соответствующего вещества в образце. В датасете для обучения присутствуют дополнительные 12 параметров, которые, собственно, и нужно научиться предсказывать: Cu, Ni, Zn, Co, Cd, Pb – по паре каждого элемента для образцов в поверхностном (0-1 см) и фоновом (15-16 см) слоях донных отложений. Приложении В приведена описательная статистика датасета для прогнозирования, в Приложении Г – датасета для обучения. В Приложении Д представлена матрица корреляций датасета для обучения (с выбросами). Заметна сильная корреляция между донными и водными показателями, а именно: между Cu, Ni, Co, Cd в верхних слоях донных отложений, Ni в нижних – и электропроводностью, Na, SO4, Cl и минерализацией водной массы.

Для наглядности посмотрим на совместное распределение параметров с высокой корреляцией: Cu в поверхностных слоях и Na в воде (рисунок 1). Заметно, что основная часть наблюдений по обоим параметрам приходится на сравнительно малые значения, однако высокая корреляция связана с чётко выраженным выбросом.

Выбросы, которые находятся от центра распределения дальше, чем на 4 сигмы, исключим из рассматриваемого набора данных. Для полученного датасета без выбросов также визуализируем матрицу корреляций, она приведена в Приложении Е. Заметно значительное снижение корреляции между донными и водными показателями.

Вернёмся к примеру с Cu в поверхностных слоях и Na в воде, совместное распределение которых, уже без выбросов, представлено на рисунке 2. Заметно отсутствие корреляции между рассматриваемыми параметрами.

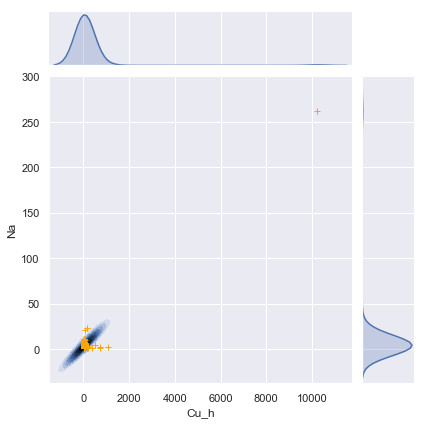


Рисунок 1. Совместное распределение Cu в поверхностных слоях и Na (с выбросом)

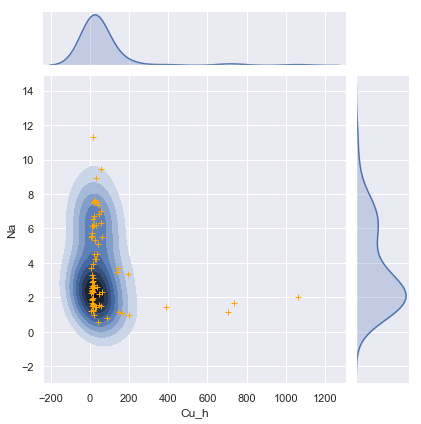


Рисунок 2. Совместное распределение Cu в поверхностных слоях и Na (без выбросов)

Аналогичная ситуация происходит и с остальными парами донных и водных показателей, что указывает на низкую корреляцию между донными и водными параметрами озёр, согласно имеющимся данным.

# Описание применяемых методов

Для решения поставленных задач необходимо провести разведывательный факторный анализ на наборе данных по 3 озёрам Мурманской области и спрогнозировать донные характеристики по водным параметрам для каталогизированных озёр Арктического региона. Применим такие методы, как разведывательный факторный анализ и метод опорных векторов.

## Разведывательный факторный анализ

Цель разведывательного факторного анализа состоит в обнаружении латентных, то есть скрытых, переменных, в результате влияния которых появляется ковариация среди наблюдаемых переменных. На этапе выделения факторов эта совместная вариация переменных отделяется от собственной вариации переменной, чтобы показать скрытую структуру факторов. Так, на результате отражается только совместная вариация, то есть латентные факторы.

В отличие от факторного анализа, метод главных компонент РСА не делает различий между совместной и собственной вариацией переменных, что может привести к ошибочным результатам (Anna B. Costello, 2006).

Для определения достаточного количества факторов оценивается значение собственного числа каждого фактора. Общая рекомендация состоит в том, чтобы принимать факторы, собственное число которых превосходит 1. Это означает, что данный фактор объясняет долю вариации больше, чем одна любая рассматриваемая переменная. Среди других рекомендаций по выбору количества факторов стоит отметить следующие: выбирать такое количество, которое обеспечивает объяснение заданного достаточного уровня вариации (например, более 70%); руководствоваться графиком собственных значений или доли объяснённой вариации (scree plot), выбирая то количество факторов, дальше которых кривая становится плоской .

Объём данных, необходимый для корректного проведения анализа, рекомендуется устанавливать из соотношения количества параметров к количеству записей 10:1. Однако большая часть исследований проводится при соотношении 2:1 и меньше (Anna B. Costello, 2006).

Результатом факторного анализа является полученный ряд факторов, каждый из которых характеризуется той или иной степенью вхождения в него начальных переменных, то есть нагрузкой (factor loadings).

Для проведения факторного анализа воспользуемся пакетом psych для R, в котором реализован разведывательный факторный анализ.

## Метод опорных векторов

Как правило, метод опорных векторов (SVM) применяется для задач классификации, однако он так же может быть использован для решения задач регрессии (M. O. Stitson, 1996). Его фундаментальный принцип состоит в нахождении и максимизации зазора между разделяемыми классами путём построения опорных векторов.

При решении задач регрессии принимается предположение, что значение целевой переменной может задаваться как линейная комбинация неких порождающих функций, которые зависят только от предикторов.

Для построения квантильной регрессии строятся, соответственно, столько линейных комбинаций для каждой целевой переменной, сколько задано интересующих квантилей распределения. Полученные результаты можно трактовать как доверительные интервалы оценок распределения.

Также имеются модификации данного метода, например взвешенная квантильная регрессия методом опорных векторов (Qifa Xu, 2015) – Support Vector Weighted Quantile Regression (SVWQR), в котором реализовывается возможность решать задачи регрессии для нелинейных функций.

Квантильная регрессия методом SVM реализована в библиотеке liquidSVM для Python.

# Анализ данных и результаты

Как было описано выше, на датасете по 3 озёрам Мурманской области проведём разведывательный факторный анализ с помощью пакета psych для R, а на каталогизированном датасете по озёрам Арктического региона проведём прогнозирование донных отложений по водным характеристикам методом опорных векторов.

## Факторный анализ 3 озёр Мурманской области

Факторный анализ был проведён для 4 разных датасетов: 3 по каждому озеру в частности и 1 общий для всех трёх озёр. Получившиеся scree plot, на которых отмечены собственные значения для всех факторов и доля объяснённой вариации для первых 10 факторов, входящих в анализ, представлены на рисунках 3-6. На основании таких графиков принимается решение о достаточном количестве факторов.

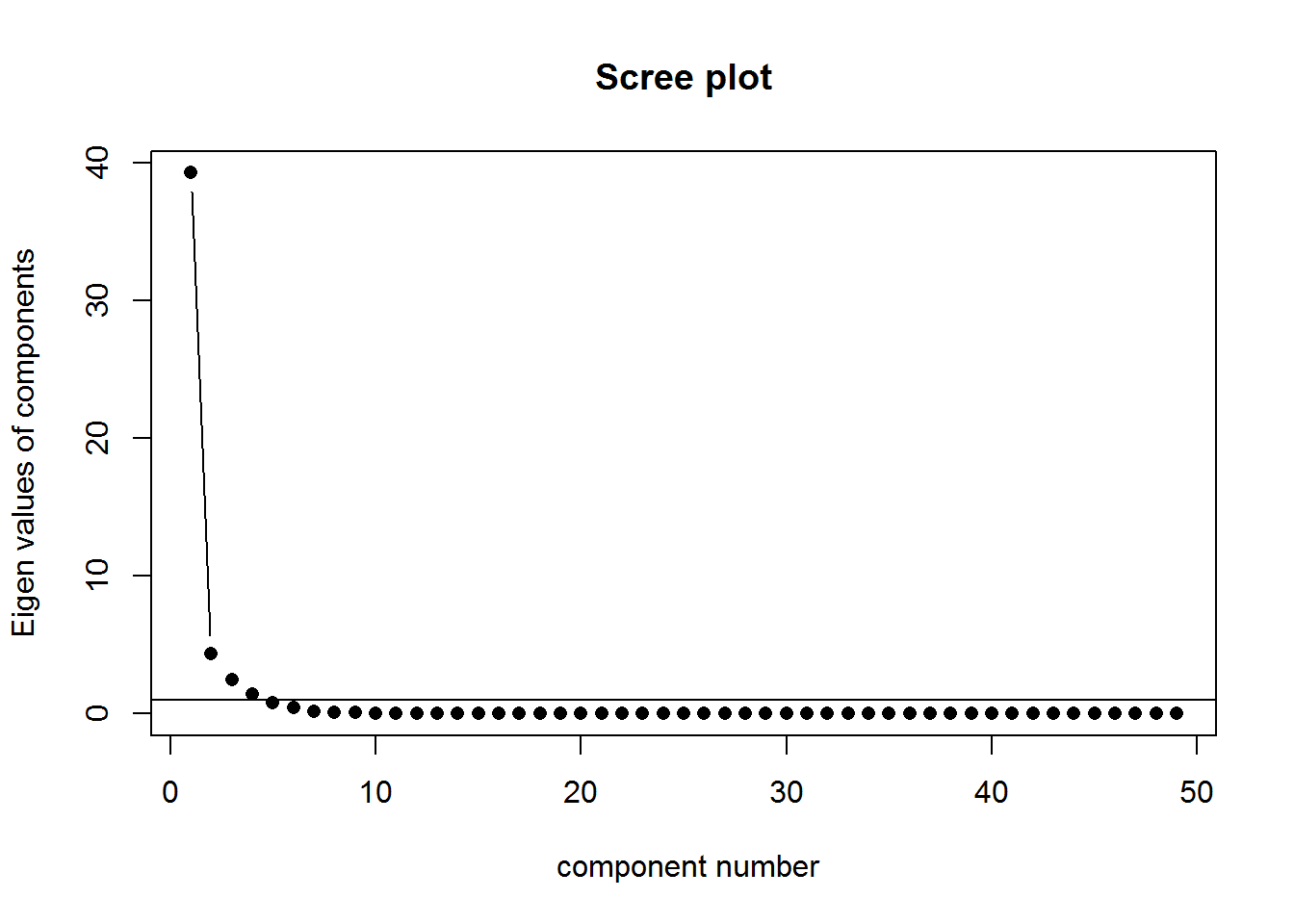
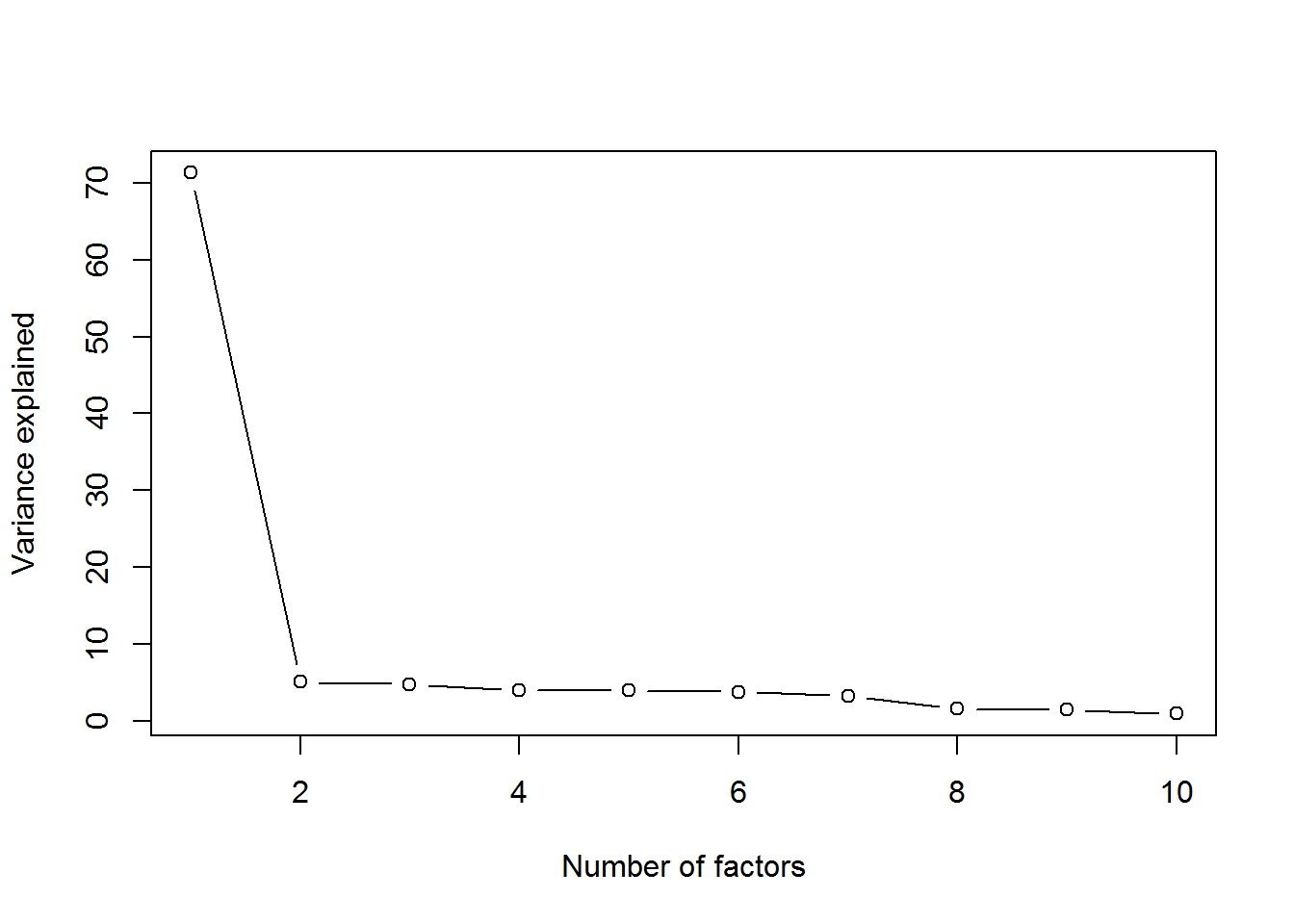
 

Рисунок 3. Scree plot: собственные значения и доля объяснённой вариации по 1 озеру

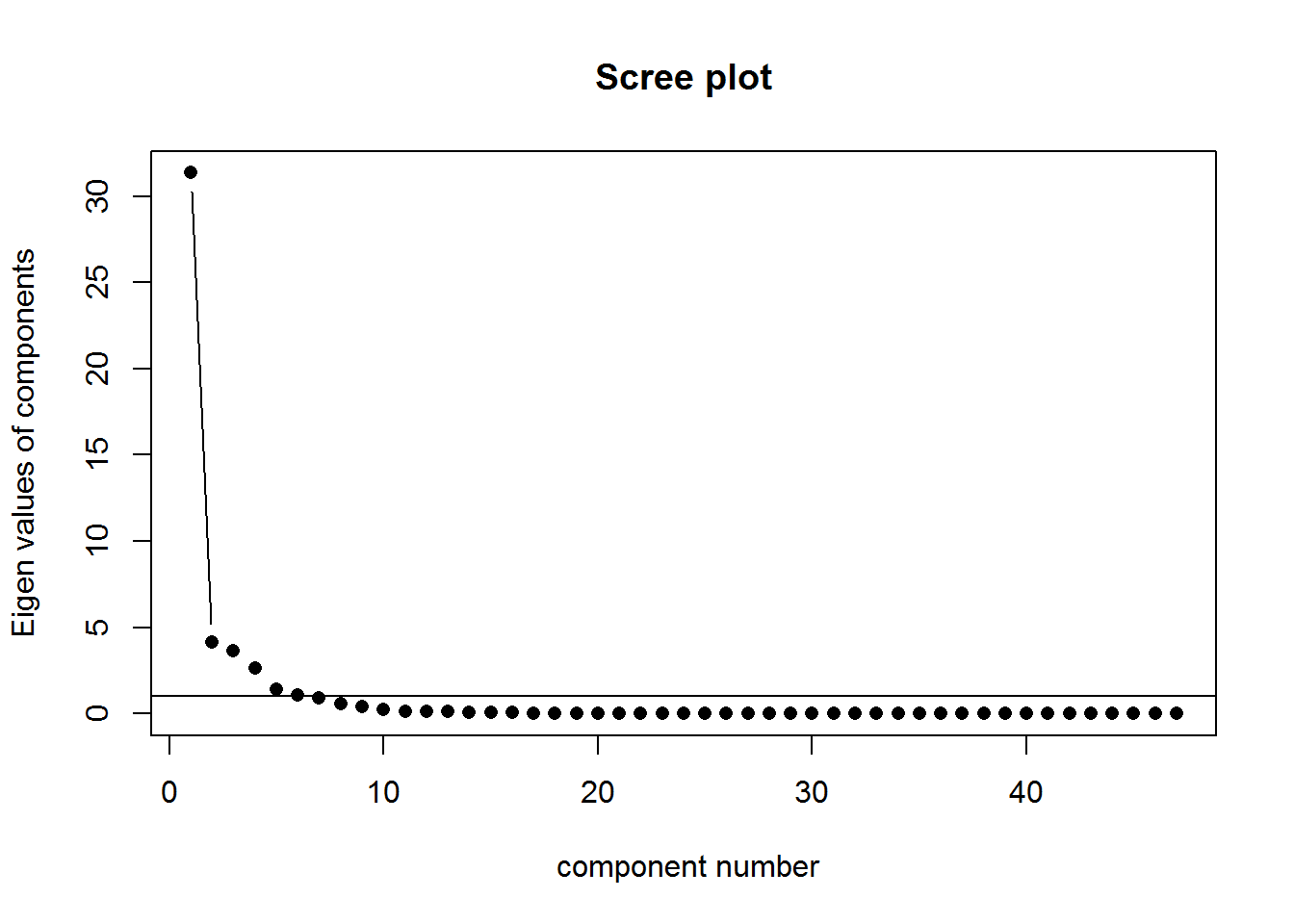
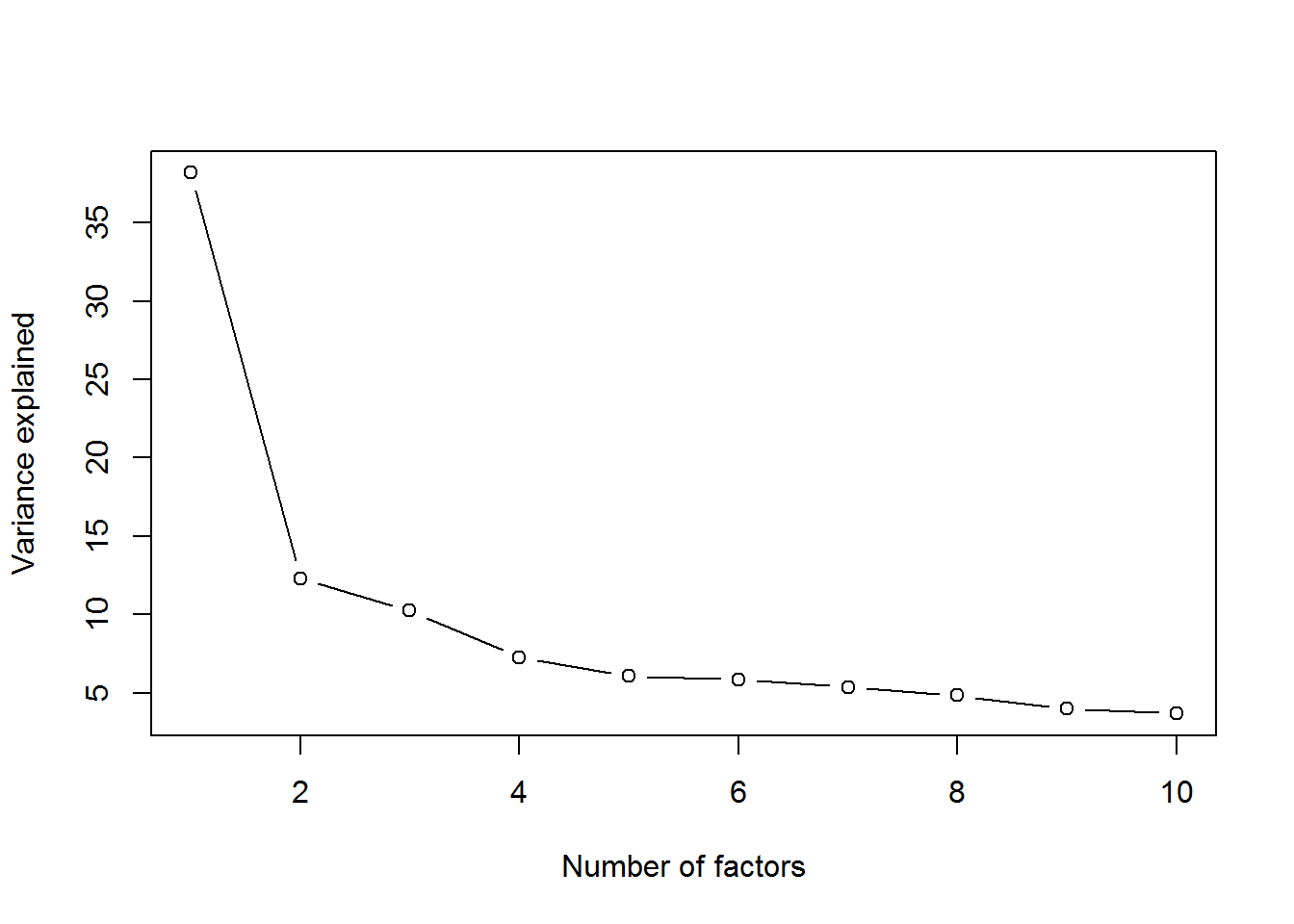
 

Рисунок 4. Scree plot: собственные значения и доля объяснённой вариации по 2 озеру

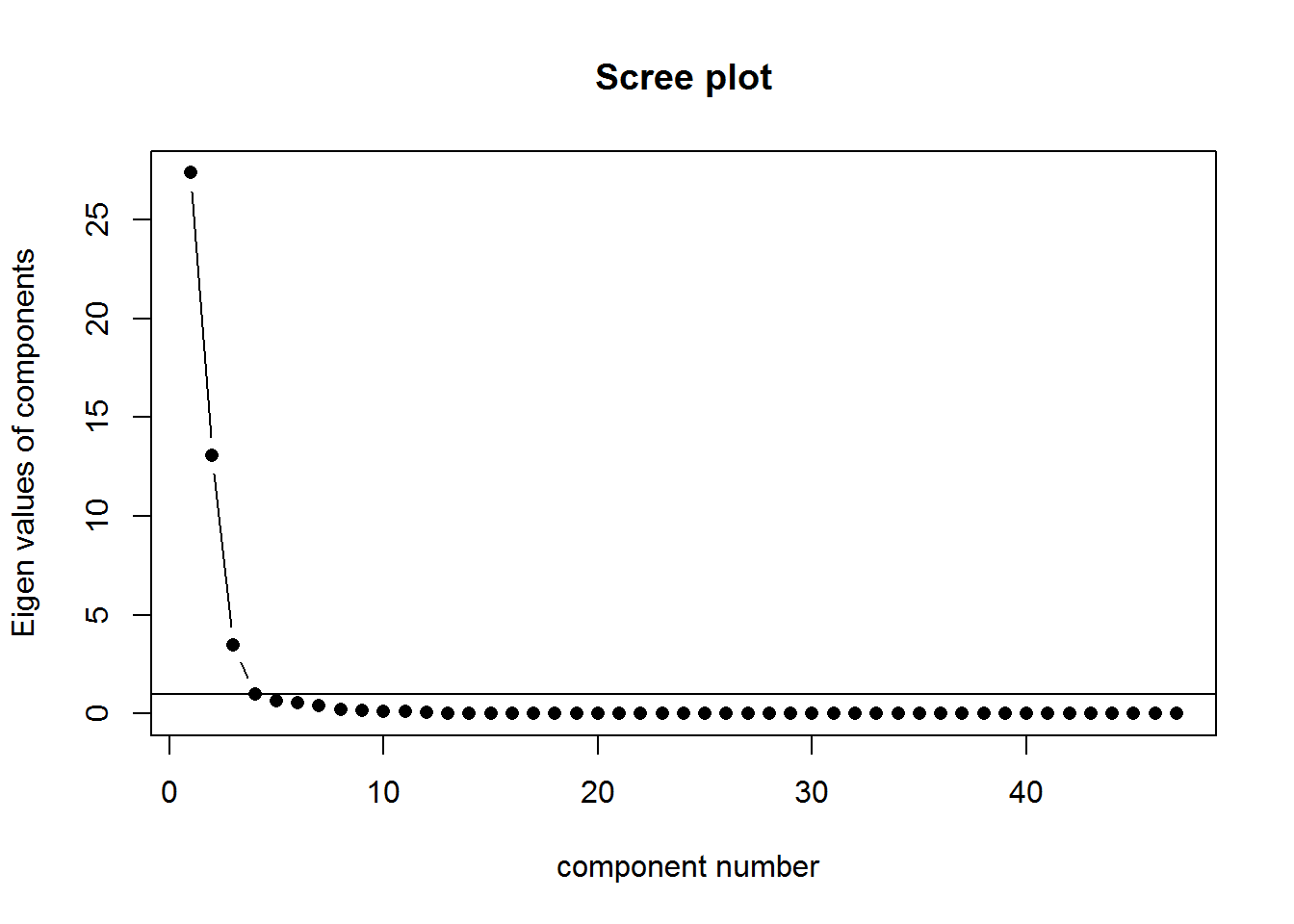
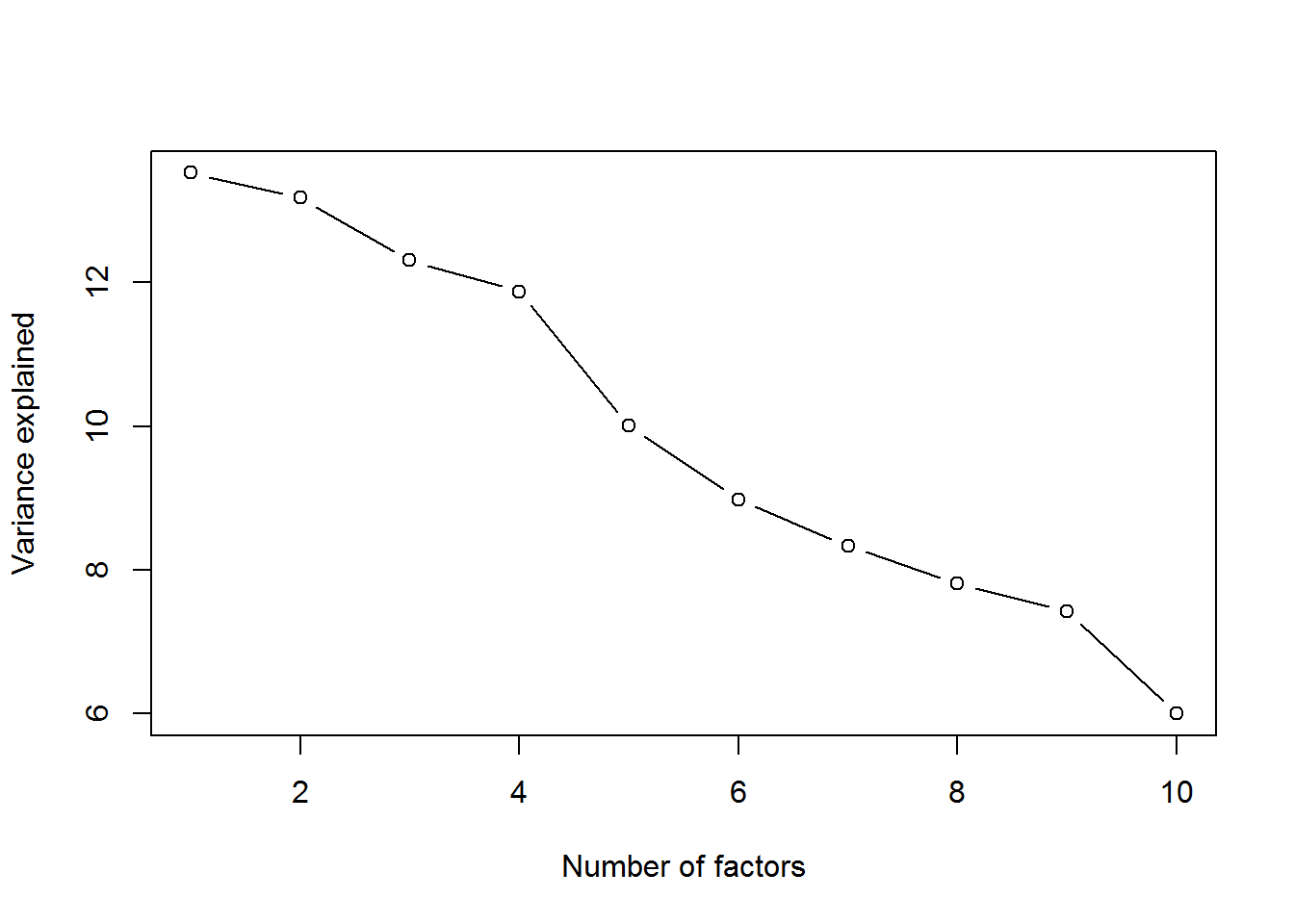
 

Рисунок 5. Scree plot: собственные значения и доля объяснённой вариации по 3 озеру

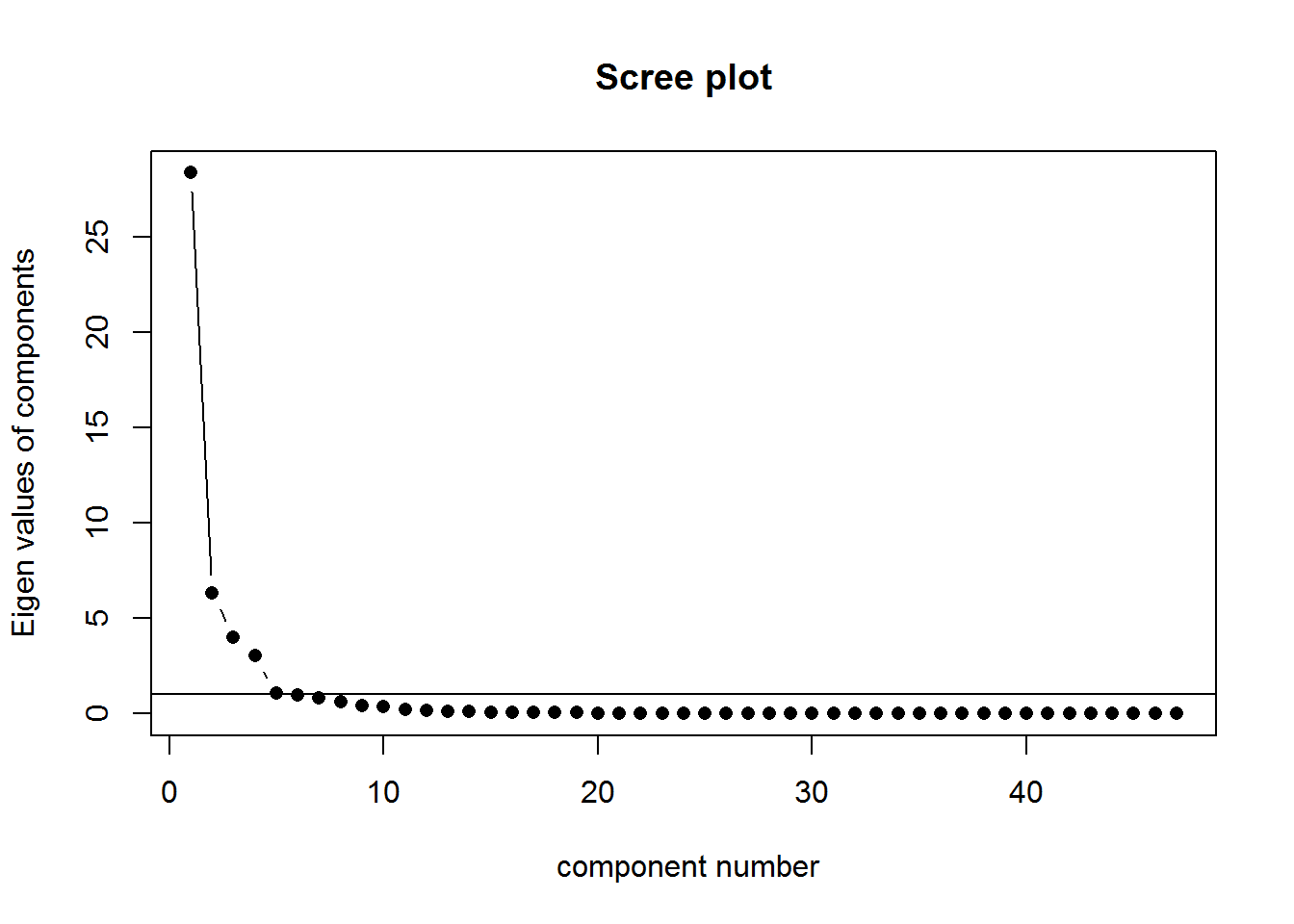
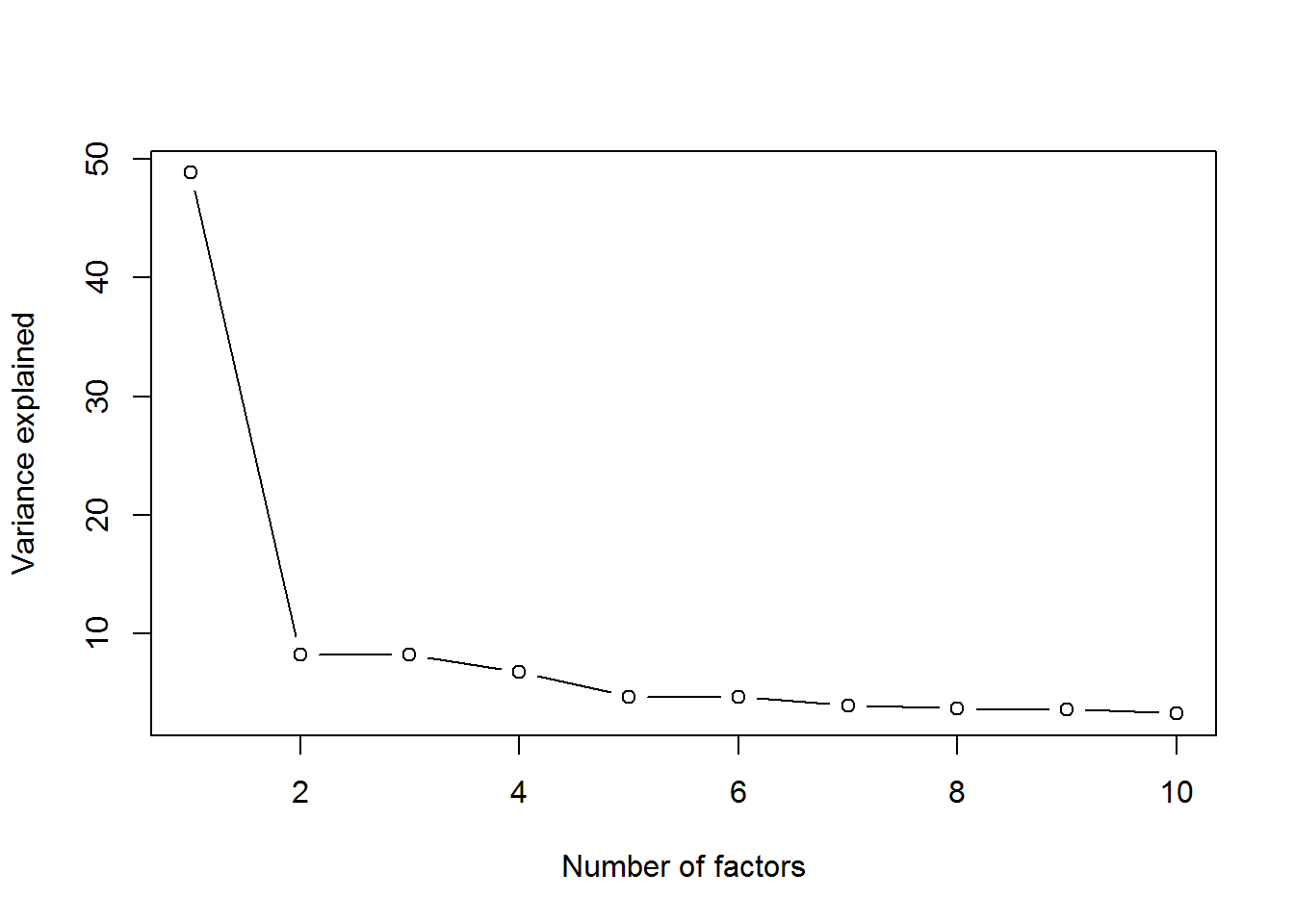
 

Рисунок 6. Scree plot: собственные значения и доля объяснённой вариации суммарно

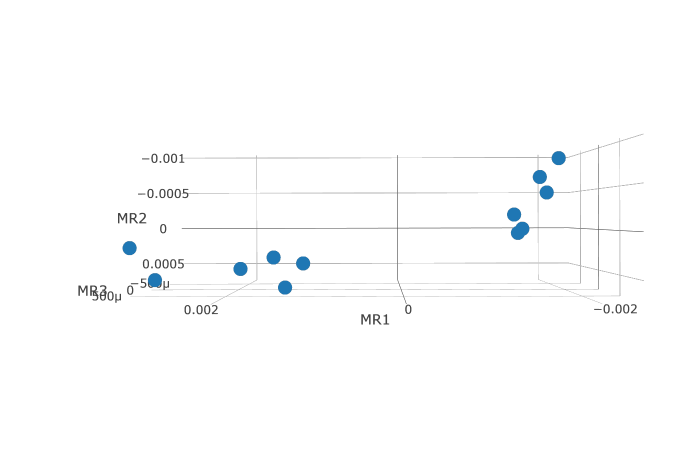
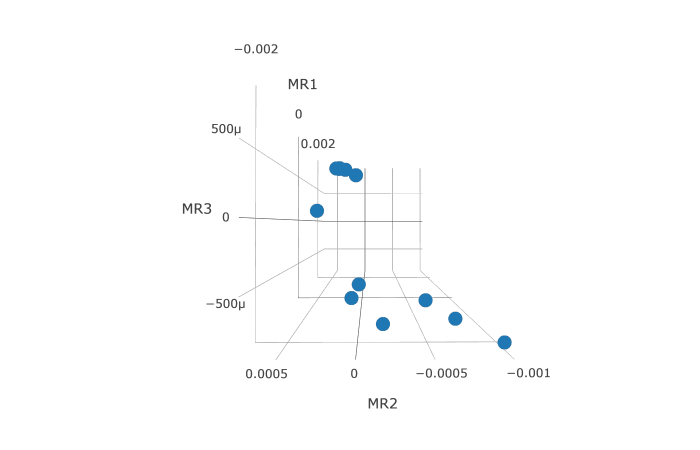
Таким образом, для 1 озера взяты 3 фактора для анализа, для 2 озера анализ проведён с 2 и 3 факторами, для 3 озёра – 3 фактора, и в совмещённом случае рассмотрены 2, 3 и 4 фактора. В Приложении Ж приведены факторные нагрузки для каждого рассмотренного случая. В таблице 1 приведена основная характеристика выделенных факторов, а именно: SS loadings – суммарная нагрузка фактора, Proportion Var – доля объяснённой вариации данным фактором, Cumulative Var – накопленная доля объяснённой вариации данным и предыдущими факторами.

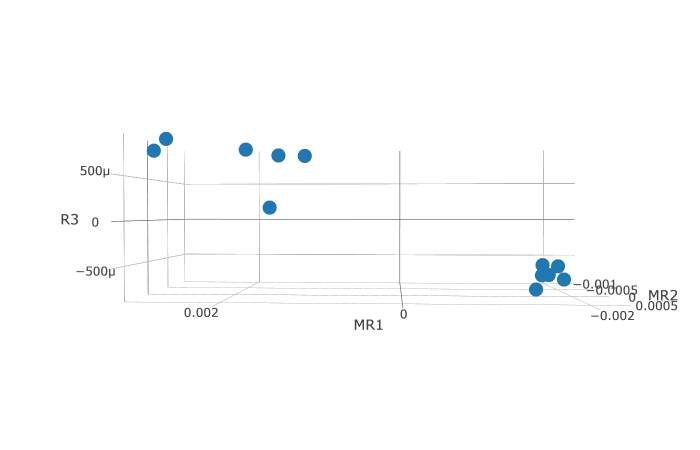
Таблица 1. Характеристика выделенных факторов

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Озеро 1 | | | Озеро 2 | | Озеро 2 | | | Озеро 3 | | |
| MR1 | MR2 | MR3 | MR1 | MR2 | MR1 | MR2 | MR3 | MR1 | MR2 | MR3 |
| SS loadings | 37.66 | 4.307 | 2.613 | 26.9 | 5.779 | 27.21 | 5.569 | 3.360 | 24.05 | 13.51 | 4.330 |
| Proportion Var | 0.769 | 0.088 | 0.053 | 0.574 | 0.123 | 0.579 | 0.118 | 0.071 | 0.512 | 0.287 | 0.092 |
| Cumulative Var | 0.769 | 0.856 | 0.910 | 0.574 | 0.697 | 0.579 | 0.697 | 0.769 | 0.512 | 0.799 | 0.891 |

Получены следующие результаты факторного анализа. Для 1 озера было выделено 3 фактора, суммарно они объясняют вариацию данных 91%. Для 2 озера было выделено 2 и 3 фактора, в каждом случае латентные переменные объясняют 69.7% и 76.9% вариации, рекомендуется сделать выбор в пользу 2 факторов, потому что такая модель лучше интерпретируется. Для 3 озера было выделено 3 фактора, которые суммарно объясняют 89,1% вариации данных.

Построим визуализацию данных каждого озера в полученном пространстве факторов (рисунки 7-9). Так, в зависимости от количества выделенных факторов, все имеющиеся данные трансформируются в 2-х или 3-х размерное пространство. На примере 1 озера ясно видно, что наблюдения делятся на 2 кластера в зависимости от значений, принимаемых фактором MR1.



М

Рисунок 7. 1 озеро в пространстве 3 факторов

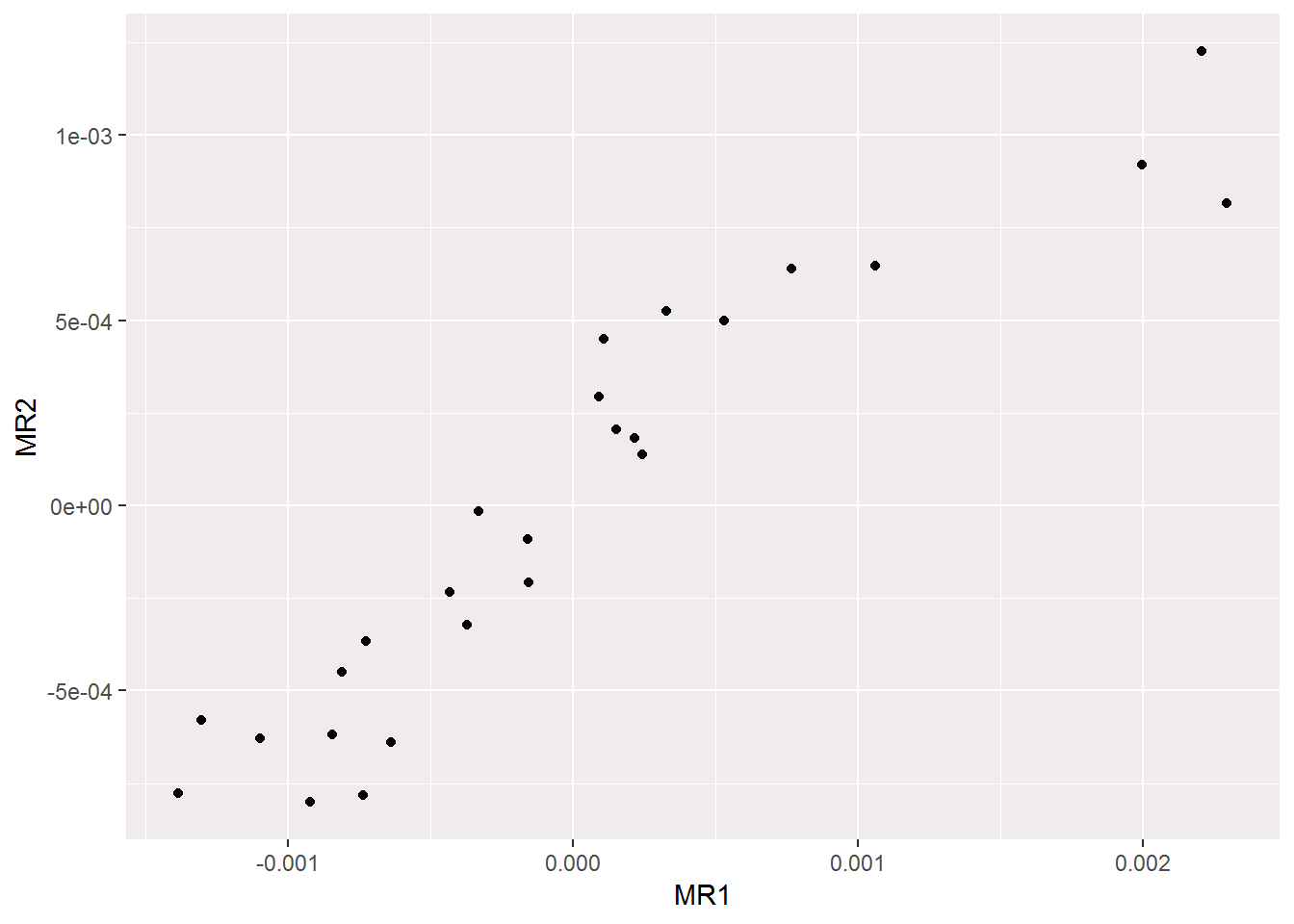
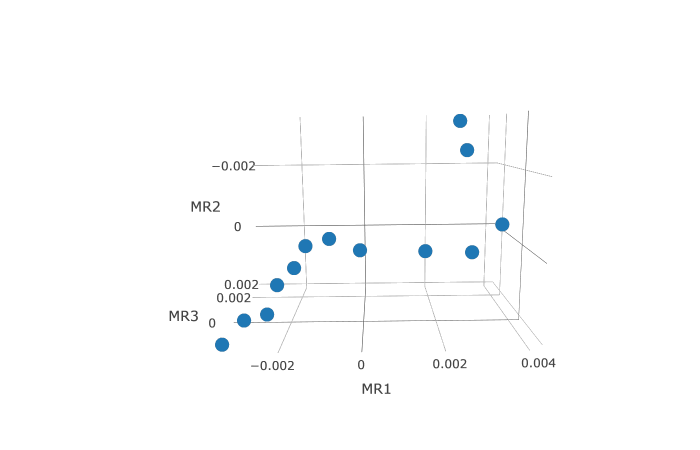
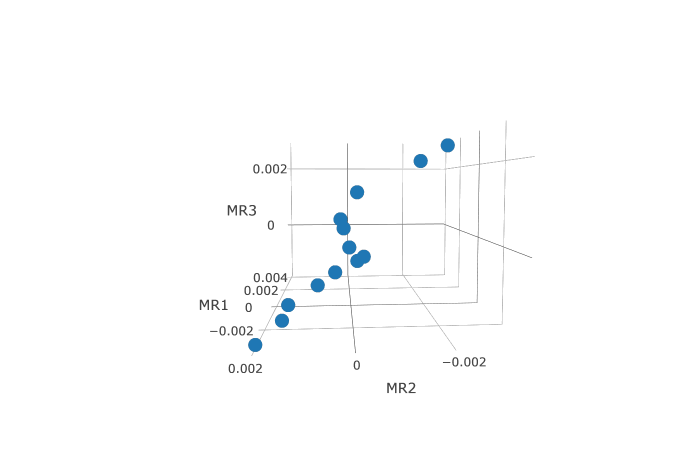


Рисунок 8. 2 озеро в пространстве 2 факторов

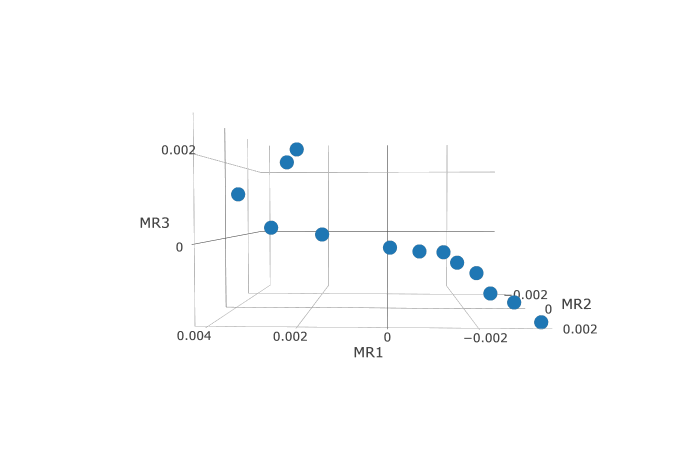


Рисунок 9. 3 озеро в пространстве 3 факторов

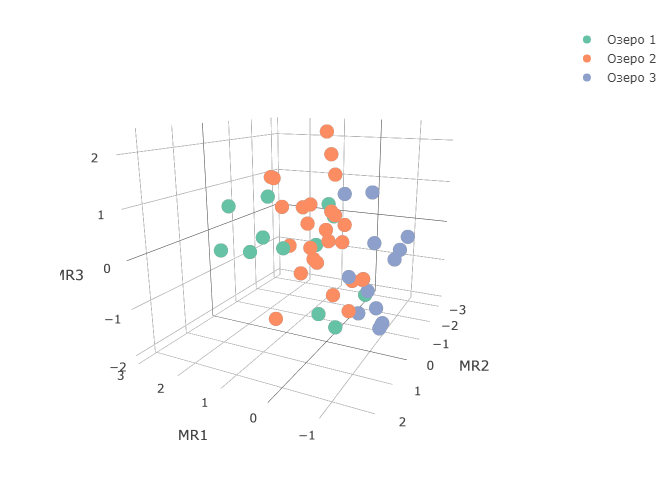
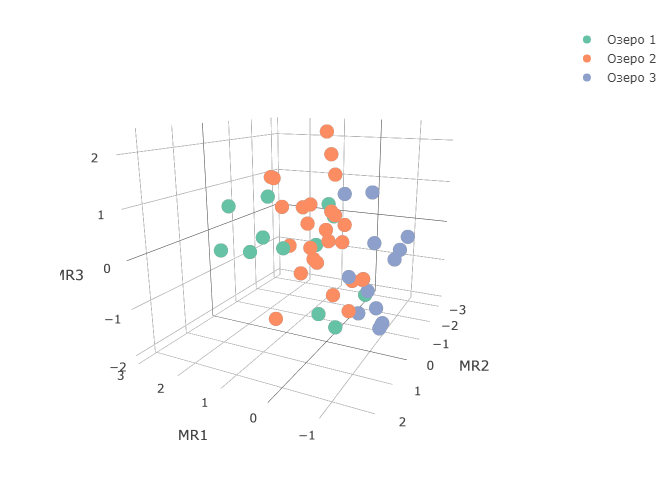
Вернёмся к совмещённому по всем трём озёрам датасету. Как было отмечено ранее, в совмещённом случае рассмотрены 2, 3 и 4 фактора. Представим полученную характеристику в Таблице 2.

Таблица 2. Характеристика выделенных факторов для совмещённого датасета

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | MR1 | MR2 | MR1 | MR2 | MR3 | MR1 | MR2 | MR3 | MR4 |
| SS | 28.097 | 6.034 | 26.451 | 5.397 | 5.340 | 26.198 | 5.269 | 5.115 | 3.224 |
| Proportion Var | 0.598 | 0.128 | 0.563 | 0.115 | 0.114 | 0.557 | 0.112 | 0.109 | 0.069 |
| Cumulative Var | 0.598 | 0.726 | 0.563 | 0.678 | 0.791 | 0.557 | 0.670 | 0.778 | 0.847 |

При выделении 2 факторов доля накопленная доля объяснённой вариации равна 72.6%, при выделении 3 факторов она повышается до 79.1%, при выделении 4 – до 84.7%. Рекомендуется рассматривать 3 фактора как оптимальный вариант, потому что добавление дополнительного фактора объясняет всего 5% вариации, однако интерпретация результатов при этом может усложниться.

Визуализируем полученную модель на рисунке 10.

М

Рисунок 10. Совмещённый датасет в пространстве 3 факторов

## Прогнозирование геохимических характеристик озёр Арктического региона

Для прогнозирования воспользуемся методом опорных векторов SVM, при этом будем прогнозировать квантили распределения. Квантильная регрессия методом SVM реализована в библиотеке liquidSVM для Python.

Выберем квантили: 0.05, 0.25, 0.5, 0.75 и 0.95. Задача прогнозирования звучит следующим образом: предсказать квантили распределения донных элементов (Cu\_h, Ni\_h, Zn\_h, Co\_h, Cd\_h, Pb\_h и Cu\_l, Ni\_l, Zn\_l, Co\_l, Cd\_l, Pb\_l) по характеристикам воды (pH, electricity, mineralization, alkalinity, colour, Ca, Mg, Na, K, HCO3, SO4, Cl, NH4, NO3, N, PO4, P, Fe). Разделим выборку на обучающую, на основе которой будет тренировать модель, и тестовую, по которой будем оценивать качество модели. Для каждого химического элемента будет построена собственная регрессионная модель по 5 рассматриваемым квантилям.

Результаты будем оценивать по нескольким метрикам: RMSE и r^2 относительно медианных значений (квантиль 0.5), Accuracy для двух доверительных интервалов (от квантиля 0.25 до 0.75, и от квантиля 0.05 до 0.95). Расчёт RMSE и r^2 реализован в библиотеке sklearn.metrics, а для расчёта Accuracy воспользуемся формулой:

, где – рассматриваемые квантили распределения, и – значение переменной, предсказанной для данных квантилей, – истинное значение предсказываемой переменной.

По своей сути Accuracy показывает долю истинных значений, попадающих в предсказанный доверительный интервал. То есть то, насколько верно модель строит доверительный интервал для прогнозируемого параметра.

Полученные значения метрик представим в таблице 3.

Таблица 3. Результаты квантильной регрессии методом SVM

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Cu\_h | Ni\_h | Zn\_h | Co\_h | Cd\_h | Pb\_h | Cu\_l | Ni\_l | Zn\_l | Co\_l | Cd\_l | Pb\_l |
| Accuracy  (0.25, 0.75) | 0.52 | 0.52 | 0.52 | 0.26 | 0.57 | 0.3 | 0.48 | 0.39 | 0.48 | 0.35 | 0.7 | 0.22 |
| Accuracy  (0.05, 0.95) | 0.91 | 0.91 | 0.87 | 0.78 | 0.91 | 0.83 | 0.74 | 0.83 | 0.87 | 0.78 | 0.91 | 0.7 |
| RMSE | 2 132.6 | 9 964.78 | 65.79 | 139.09 | 2.98 | 25.84 | 22.3 | 94.87 | 79.95 | 12.21 | 0.41 | 8.49 |
| r^2 | -0.05 | -0.05 | -0.04 | -0.09 | -0.08 | -0.17 | -0.21 | -0.08 | -0.11 | -0.12 | -0.18 | -0.06 |

По полученным метрикам можно сказать, что полученные модели очень плохо предсказывают концентрацию донных элементов по водным характеристикам. Однако 95%-ый доверительный интервал был предсказан достаточно корректно. Величина r^2 получилась отрицательной, что указывает на отсутствие какой-либо взаимосвязи между предикторами и целевыми переменными, согласно имеющимся данным.

В дальнейшем рекомендуется расширять набор данных для получения более точных оценок и выводов. В имеющихся данных наблюдения относятся к целевым переменным как 2:1, чего может быть достаточно для факторного анализа, но не хватает для более сложных методов машинного обучения.

Также при использовании SVM для регрессии делается предположение о линейной зависимости целевых переменных и предикторов, которое может быть не обосновано на имеющихся данных. Рекомендуется использовать иные методы и алгоритмы для прогнозирования, которые не принимают подобного предположения.

# Заключение

В данной работе была предпринята попытка применить методы машинного обучения для решения задач, стоящих перед наукой геохимией.

Был проведён факторный анализ химических элементов донных отложений 3 озёр Мурманской области, в результате которого были выделены 3, 2 и 3 фактора для каждого озера соответственно. Был также проведён факторный анализ по совмещённым для всех трёх озёр наблюдениям, в результате которого было выделено 3 фактора. Доля объяснённой вариации в каждом из 4 приведённых случаев равна, соответственно, 91%, 69.7%, 89.1%, 79.1%. Стоит отметить, что факторный анализ является одним из традиционных методов, применяющихся для нужд геохимии.

Была построена квантильная регрессия методом опорных векторов для каталогизированных данных по водным характеристикам и донным отложениям озёр Арктического региона, а именно Мурманской области. Рассматривались квантили распределения 0.05, 0.25, 0.5, 0.75, 0.95. Регрессия строилась на основании параметров состава воды, которые являются предикторами для прогнозирования концентрации химических элементов в донных отложениях. Качество полученной модели было оценено по следующим метрикам: Accuracy для доверительных интервалов, RMSE и r^2 для медианных значений. Получившееся качество модели не удовлетворяет условиям, необходимым для корректного прогнозирования.

В дальнейшем рекомендуется использовать иные методы прогнозирования, а также повышать объём данных для анализа, так как текущее соотношение наблюдений к переменным как 2:1 не позволяет использовать более сложные методы.

# Список литературы

1. **Application of Artificial Neural Networks in the River Water Quality Modeling: Karoon River, Iran** [Журнал] / авт. Musavi-Jahromi S.H. Golabi M.. - [б.м.] : Journal of Applied Sciences, 2008 г.. - 12 : Т. 8. - стр. 2324-2328.
2. **Application of Factor Analysis in the Identification of a Geochemical Signature of Buried Kimberlites in Near-Surface Groundwaters in the Attawapiskat Area of the James Bay Lowlands of Northern Ontario, Canada** [Отчет] / авт. Drouin Marc. - Ottawa : Ottawa-Carleton Geoscience Centre, Faculty of Science, 2012.
3. **Artificial neural network modeling of the river water quality—A case study** [Журнал] / авт. Kunwar P.Singh Ankita Basant, Amrita Malik, Gunja Jain. - [б.м.] : Ecological Modelling, 2009 г.. - 6 : Т. 220. - стр. 888-895.
4. **Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis** [Журнал] / авт. Anna B. Costello Jason W. Osborne. - [б.м.] : Practical Assessment, Research & Evaluation, 2006 г.. - 7 : Т. 10.
5. **Factor analysis applied to regional geochemical data: problems and possibilities** [Журнал] / авт. Clemens Reimann Peter Filzmoser, Robert G.Garrett // Applied Geochemistry. - 2002 г.. - 3 : Т. 17. - стр. 185-206.
6. **Machine Learning Approach to Predict Sediment** [Журнал] / авт. Hazi Md. Azamathulla Aminuddin Ab. Ghani, Chun Kiat Chang, Zorkeflee Abu Hasan, Nor Azazi Zakaria. - 969-976 : Clean – Soil, Air, Water, 2010 г.. - 10 : Т. 38.
7. **Principal component analysis and hierarchical cluster analyses of arsenic groundwater geochemistry in the Hetao basin, Inner Mongolia** [Журнал] / авт. Yuxiao Jiang Huaming Guo, Yongfeng Jia, Yongsheng Cao, Chao Hu // Chemie der Erde. - 2015 г.. - Т. 75. - стр. 197-205.
8. **Principal component analysis of fluoride geochemistry of groundwater in Shanxi and Inner Mongolia, China** [Журнал] / авт. Shan Hu Ting Luo, Chuanyong Jing // Journal of Geochemical Exploration. - 2013 г.. - Т. 135. - стр. 124-129.
9. **Progress and challenges of big data research on petrology and geochemistry** [Журнал] / авт. Shoutao Jiao Qi Zhang, Yongzhang Zhou, Wanfeng Chen, Xinyu Liu, Gnanachandrasamy Gopalakrishnan // Solid Earth Sciences. - 2018 г.. - Т. 3. - стр. 105-114.
10. **Theory of Support Vector Machines** [Отчет] / авт. M. O. Stitson J. A. E. Weston, A. Gammerman, V. Vovk, V. Vapnik. - Egham : Royal Holloway University of London, 1996.
11. **Weighted quantile regression via support vector machine** [Журнал] / авт. Qifa Xu Jinxiu Zhang, Cuixia Jiang, Xue Huang, Yaoyao He // Expert Systems with Applications. - 2015 г.. - 13 : Т. 42. - стр. 5441-5451.
12. **Аннотированный экологический каталог озер Мурманской области: Восточная часть Мурманской области (бассейн Баренцева моря)** [Книга] / авт. Н.А. Кашулин С.С. Сандимиров, В.А. Даувальтер, Л.П. Кудрявцева, П.М. Терентьев, Д.Б. Денисов, С.А. Валькова. - Апатиты : Изд-во Кольского научного центра РАН, 2010.
13. **Аннотированный экологический каталог озер Мурманской области: центральный и юго-западный районы Мурманской области (бассейны Баренцева и Белого морей и Ботнического залива Балтийского моря)** [Книга] / авт. Н.А. Кашулин С.С. Сандимиров, В.А. Даувальтер, Л.П. Кудрявцева, П.М. Терентьев, Д.Б. Денисов, О.И. Вандыш, И.М. Королева, С.А. Валькова, Т.Г. Кашулина. - Апатиты : КНЦ РАН, 2013.
14. **Аннотированный экологический каталог озер Мурманской области: Юго-Восточная часть (бассейн Белого моря)** [Книга] / авт. Н.А. Кашулин С.С. Сандимиров, В.А. Даувальтер, Л.П. Кудрявцева, П.М. Терентьев, Д.Б. Денисов, О.И. Вандыш, С.А. Валькова. - Апатиты : Изд-во Кольского научного центра РАН, 2012.
15. **Использование математических методов в оценке состояния экосистемы озера Имандра по диатомовым комплексам донных отложений** [Конференция] / авт. Вокуева С. И. Денисов Д.Б. // Математические исследования в естественных науках. Труды XIV Всероссийской научной школы. - Апатиты : [б.н.], 2017. - стр. 134-145.
16. **Научные мероприятия. Всероссийская научная школа «Математические исследования в естественных науках** [В Интернете]. - 12 Июнь 2019 г.. - http://geoksc.apatity.ru/index.php/actions.
17. **Применение искусственных нейронных сетей для прогнозирования и визуализации пространственного распределения химических элементов в поверхностном слое почвы** [Конференция] / авт. Буевич А. Г. Бусловская А. Н., Рахматова А. Ю., Крамаренко А. А., Сергеев А. П., Тарасов Д. А // Информация: передача, обработка, восприятие: материалы международной научно-практической конференции. - Екатеринбург : УрФУ, 2016. - стр. 20-33.
18. **Применение статистических методов в исследовании химического состава донных отложений озера Имандра** [Конференция] / авт. Мицуков А. С. Даувальтер В. А. // Математические исследования в естественных науках. Труды XV Всероссийской научной школы. - Апатиты : [б.н.], 2018. - стр. 153-163.
19. **Тяжелые металлы и металлоиды в почвах** [Книга] / авт. Водяницкий Ю Н. - Москва : ГНУ Почвенный институт им. В.В. Докучаева РАСХН, 2008.

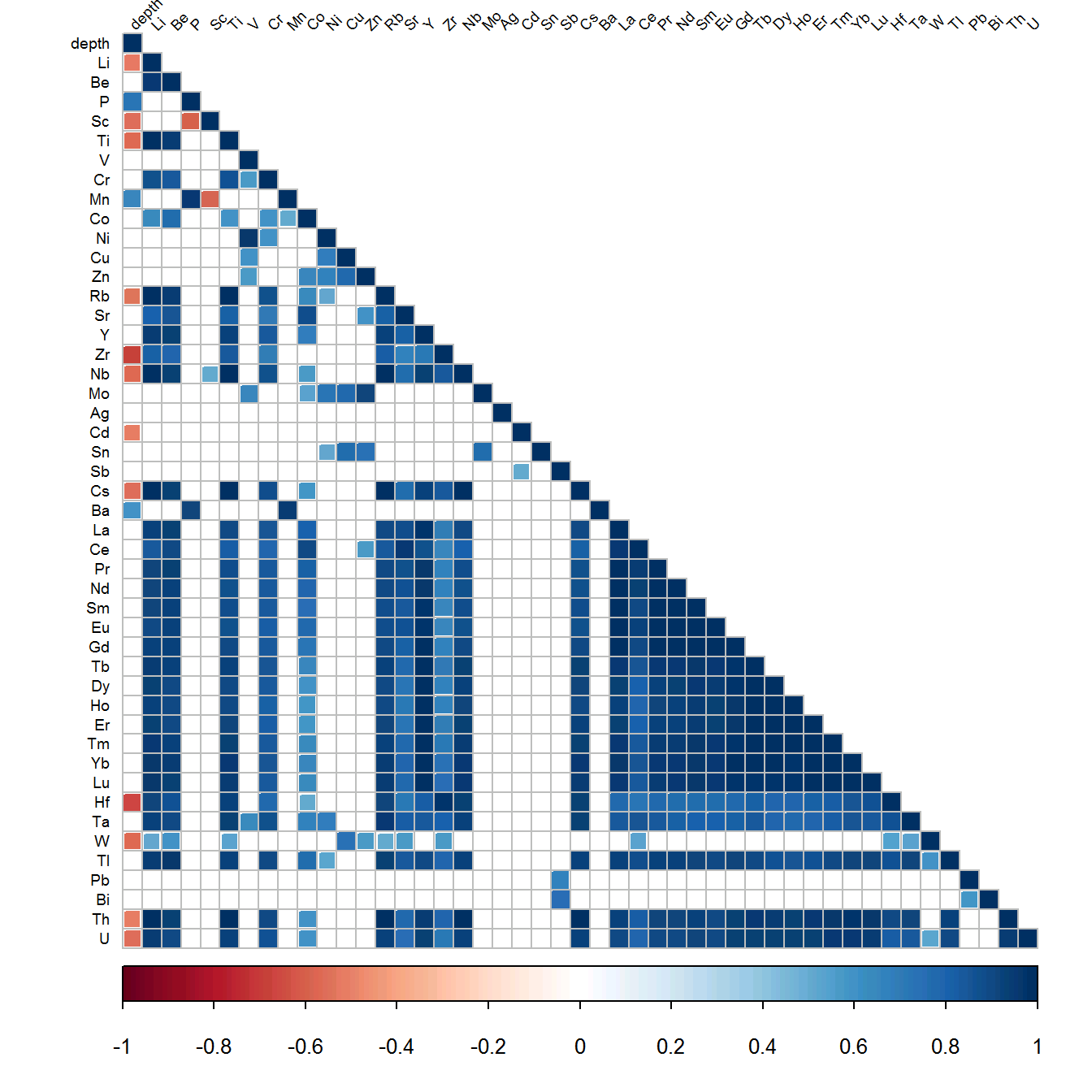
Приложение **А**

Описательная статистика данных по 3 озёрам Мурманской области.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | mean | sd | median | min | max | skew | kurtosis | se |
| depth | 17.53 | 11.81 | 15.00 | 1.00 | 47.00 | 0.70 | -0.32 | 1.65 |
| Li | 10.55 | 12.55 | 4.05 | 0.34 | 46.67 | 1.40 | 0.84 | 1.76 |
| Be | 0.68 | 0.40 | 0.59 | 0.26 | 1.84 | 1.02 | 0.21 | 0.06 |
| P | 9641.16 | 11562.99 | 2445.00 | 1058.00 | 37830.00 | 1.09 | -0.32 | 1619.14 |
| Sc | 9.99 | 4.84 | 8.87 | 2.71 | 19.49 | 0.16 | -1.37 | 0.68 |
| Ti | 1159.61 | 1149.56 | 582.00 | 113.70 | 4284.00 | 1.31 | 0.49 | 160.97 |
| V | 396.56 | 927.18 | 43.16 | 13.65 | 4785.00 | 3.34 | 11.37 | 129.83 |
| Cr | 48.86 | 45.03 | 31.67 | 10.14 | 195.80 | 1.48 | 1.54 | 6.31 |
| Mn | 1346.87 | 1267.76 | 918.20 | 65.38 | 4400.00 | 0.95 | -0.16 | 177.52 |
| Co | 13.22 | 7.68 | 14.57 | 1.81 | 26.00 | -0.27 | -1.20 | 1.08 |
| Ni | 91.65 | 137.73 | 38.07 | 9.67 | 606.50 | 2.51 | 5.73 | 19.29 |
| Cu | 229.29 | 292.72 | 111.10 | 23.64 | 1260.00 | 1.83 | 2.62 | 40.99 |
| Zn | 320.22 | 201.66 | 311.10 | 39.96 | 962.60 | 0.57 | 0.25 | 28.24 |
| Rb | 25.86 | 29.30 | 10.62 | 1.35 | 103.90 | 1.38 | 0.63 | 4.10 |
| Sr | 61.70 | 26.41 | 61.23 | 26.31 | 118.80 | 0.42 | -0.86 | 3.70 |
| Y | 8.83 | 3.88 | 7.65 | 3.98 | 20.25 | 1.33 | 1.11 | 0.54 |
| Zr | 45.48 | 31.39 | 35.94 | 6.36 | 114.10 | 0.64 | -0.94 | 4.39 |
| Nb | 2.52 | 2.46 | 1.41 | 0.23 | 9.75 | 1.43 | 1.02 | 0.34 |
| Mo | 2.95 | 2.26 | 2.68 | 0.71 | 13.32 | 2.04 | 6.47 | 0.32 |
| Ag | 0.60 | 0.28 | 0.52 | 0.22 | 1.68 | 1.38 | 2.91 | 0.04 |
| Cd | 0.74 | 0.27 | 0.72 | 0.22 | 1.32 | 0.25 | -0.75 | 0.04 |
| Sn | 6.26 | 5.88 | 3.80 | 0.38 | 22.16 | 1.11 | 0.11 | 0.82 |
| Sb | 1.15 | 0.89 | 1.12 | 0.10 | 3.90 | 0.99 | 0.72 | 0.12 |
| Cs | 1.14 | 1.21 | 0.60 | 0.05 | 4.52 | 1.41 | 0.78 | 0.17 |
| Ba | 554.68 | 560.53 | 390.10 | 77.95 | 2249.00 | 1.47 | 1.48 | 78.49 |
| La | 16.14 | 7.75 | 15.75 | 7.10 | 37.10 | 1.05 | 0.44 | 1.08 |
| Ce | 32.39 | 14.15 | 31.25 | 13.49 | 63.16 | 0.47 | -0.64 | 1.98 |
| Pr | 3.91 | 1.92 | 3.80 | 1.54 | 9.22 | 0.97 | 0.38 | 0.27 |
| Nd | 14.55 | 7.26 | 13.99 | 5.93 | 34.07 | 0.99 | 0.40 | 1.02 |
| Sm | 2.78 | 1.32 | 2.55 | 1.16 | 6.48 | 1.16 | 0.75 | 0.19 |
| Eu | 0.60 | 0.27 | 0.58 | 0.25 | 1.38 | 1.12 | 0.83 | 0.04 |
| Gd | 2.23 | 1.06 | 1.97 | 0.97 | 5.30 | 1.26 | 0.86 | 0.15 |
| Tb | 0.32 | 0.13 | 0.28 | 0.16 | 0.71 | 1.39 | 1.34 | 0.02 |
| Dy | 1.77 | 0.78 | 1.54 | 0.82 | 4.14 | 1.46 | 1.44 | 0.11 |
| Ho | 0.34 | 0.15 | 0.32 | 0.16 | 0.80 | 1.46 | 1.48 | 0.02 |
| Er | 0.98 | 0.46 | 0.87 | 0.45 | 2.38 | 1.55 | 1.67 | 0.06 |
| Tm | 0.13 | 0.07 | 0.11 | 0.06 | 0.33 | 1.40 | 1.02 | 0.01 |
| Yb | 0.85 | 0.44 | 0.69 | 0.39 | 2.13 | 1.44 | 1.10 | 0.06 |
| Lu | 0.12 | 0.07 | 0.10 | 0.05 | 0.33 | 1.64 | 2.05 | 0.01 |
| Hf | 1.21 | 0.88 | 0.85 | 0.18 | 3.52 | 0.94 | -0.12 | 0.12 |
| Ta | 0.24 | 0.26 | 0.10 | 0.02 | 1.05 | 1.45 | 1.31 | 0.04 |
| W | 1.09 | 0.88 | 0.92 | 0.15 | 4.37 | 1.39 | 2.36 | 0.12 |
| Tl | 0.30 | 0.20 | 0.24 | 0.05 | 0.80 | 0.79 | -0.36 | 0.03 |
| Pb | 45.41 | 26.78 | 44.29 | 3.19 | 137.07 | 0.85 | 1.36 | 3.75 |
| Bi | 0.25 | 0.16 | 0.24 | 0.04 | 0.98 | 2.09 | 7.05 | 0.02 |
| Th | 3.51 | 3.02 | 2.01 | 0.90 | 12.09 | 1.48 | 0.97 | 0.42 |
| U | 1.28 | 1.11 | 0.60 | 0.28 | 3.92 | 1.17 | 0.09 | 0.16 |

**Приложение Б**

Матрица корреляций для данных по 3 озёрам Мурманской области.



**Приложение В**

Описательная статистика датасета для прогнозирования по озёрам Арктического региона.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | count | mean | std | min | 25% | 50% | 75% | max |
| pH | 428 | 6,33 | 0,75 | 4,15 | 5,97 | 6,49 | 6,85 | 7,96 |
| electricity | 428 | 33,1 | 21,6 | 8 | 21 | 29 | 38 | 170 |
| Ca | 428 | 2,23 | 2,72 | 0,14 | 0,94 | 1,68 | 2,78 | 35,7 |
| Mg | 428 | 0,84 | 0,64 | 0,03 | 0,47 | 0,74 | 1 | 5,35 |
| Na | 428 | 2,81 | 2,38 | 0,31 | 1,61 | 2,18 | 3,22 | 21,9 |
| K | 428 | 0,52 | 0,64 | 0,02 | 0,23 | 0,36 | 0,58 | 7,02 |
| HCO3 | 428 | 9,06 | 10,15 | 0 | 2,6 | 6,77 | 12,12 | 102,2 |
| SO4 | 428 | 2,88 | 3,67 | 0,3 | 1,27 | 2 | 3,08 | 42 |
| Cl | 428 | 2,85 | 3,15 | 0,3 | 1,25 | 1,92 | 3,24 | 40,1 |
| mineralisation | 428 | 21,12 | 17,19 | 1,9 | 11,05 | 17,7 | 24,63 | 154,3 |
| alkalinity | 428 | 148,49 | 166,39 | 0 | 43 | 111 | 199,25 | 1675 |
| colour | 428 | 103,68 | 116,01 | 0 | 29,75 | 64 | 130 | 700 |
| NH4 | 428 | 25,58 | 72,75 | 0 | 6 | 12 | 21 | 1018 |
| NO3 | 428 | 38,11 | 221,18 | 0 | 1 | 2 | 8 | 2744 |
| N | 428 | 325,17 | 311,87 | 65 | 180 | 240 | 367 | 2936 |
| PO4 | 428 | 2,03 | 4,93 | 0 | 0 | 1 | 2 | 44 |
| P | 428 | 12,13 | 13,69 | 1 | 5 | 8 | 14 | 131 |
| Fe | 428 | 224,58 | 350,33 | 3 | 41 | 100 | 240 | 2700 |

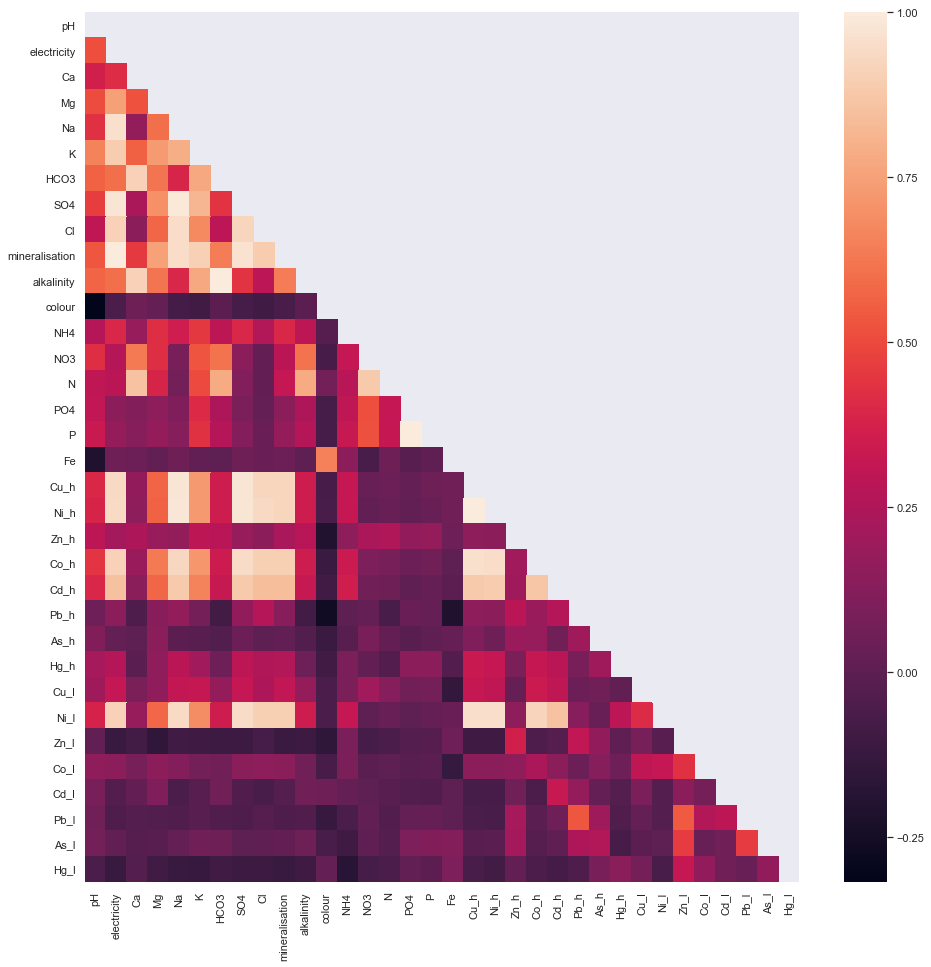
**Приложение Г**

Описательная статистика датасета для обучения по озёрам Арктического региона.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | count | mean | std | min | 25% | 50% | 75% | max |
| pH | 92 | 6,72 | 0,6 | 4,64 | 6,5 | 6,74 | 6,99 | 8,85 |
| electricity | 92 | 60,74 | 138,28 | 8 | 24 | 36 | 54,25 | 1296 |
| Ca | 92 | 3,32 | 6,77 | 0,15 | 0,96 | 2,04 | 2,86 | 60 |
| Mg | 92 | 1,16 | 1,57 | 0,01 | 0,44 | 0,82 | 1,29 | 10,9 |
| Na | 92 | 7,17 | 27,19 | 0,56 | 1,88 | 2,78 | 6,14 | 262,1 |
| K | 92 | 1,23 | 2,28 | 0,07 | 0,34 | 0,56 | 0,92 | 17 |
| HCO3 | 92 | 15,32 | 21,24 | 0,06 | 5,27 | 9,75 | 17,85 | 166,5 |
| SO4 | 92 | 9,42 | 44,09 | 0,4 | 1,67 | 2,45 | 4,2 | 418,2 |
| Cl | 92 | 4,52 | 12 | 0,6 | 1,1 | 1,7 | 2,92 | 111,5 |
| mineralisation | 92 | 41,98 | 98,19 | 4 | 15,85 | 24,95 | 35,8 | 908,5 |
| alkalinity | 92 | 251,09 | 348,18 | 1 | 86,5 | 160 | 291,75 | 2728 |
| colour | 92 | 30,24 | 35,37 | 3 | 10 | 18 | 38,25 | 245 |
| NH4 | 92 | 24,09 | 31,94 | 1 | 7 | 14,5 | 24 | 188 |
| NO3 | 92 | 171,61 | 584,28 | 1 | 2,75 | 14 | 55,5 | 3507 |
| N | 92 | 511,79 | 1065,05 | 59 | 149,25 | 248,5 | 425 | 8107 |
| PO4 | 92 | 7,98 | 48,19 | 0 | 1 | 1 | 2 | 445 |
| P | 92 | 19,65 | 72,18 | 2 | 5 | 8 | 13,25 | 670 |
| Fe | 92 | 137,17 | 210,98 | 6 | 25 | 61,5 | 124,5 | 1260 |
| Cu\_h | 92 | 185,9 | 1071,33 | 6,1 | 14,92 | 23,7 | 53 | 10234 |
| Ni\_h | 92 | 663,91 | 4982,25 | 11 | 22 | 36,5 | 75,25 | 47805 |
| Zn\_h | 92 | 105,89 | 73,28 | 11 | 58 | 97,5 | 132,25 | 487 |
| Co\_h | 92 | 25,23 | 71,62 | 1,9 | 5 | 8,2 | 23,7 | 661 |
| Cd\_h | 92 | 0,8 | 1,61 | 0,02 | 0,2 | 0,39 | 0,74 | 14,4 |
| Pb\_h | 92 | 20,17 | 23,01 | 1,5 | 6,15 | 12,5 | 25,2 | 135,6 |
| Cu\_l | 92 | 21,79 | 16,32 | 2,4 | 11,5 | 18 | 27,52 | 94 |
| Ni\_l | 92 | 27,57 | 48,47 | 5 | 12,3 | 20 | 27,25 | 468 |
| Zn\_l | 92 | 82,99 | 59,54 | 16 | 41 | 73 | 100,25 | 390 |
| Co\_l | 92 | 10,06 | 7,93 | 1,3 | 4,38 | 8,95 | 12 | 53 |
| Cd\_l | 92 | 0,32 | 0,41 | 0,01 | 0,08 | 0,17 | 0,3 | 2,1 |
| Pb\_l | 92 | 4,55 | 4,99 | 0,4 | 1,58 | 3,51 | 5,08 | 35,1 |

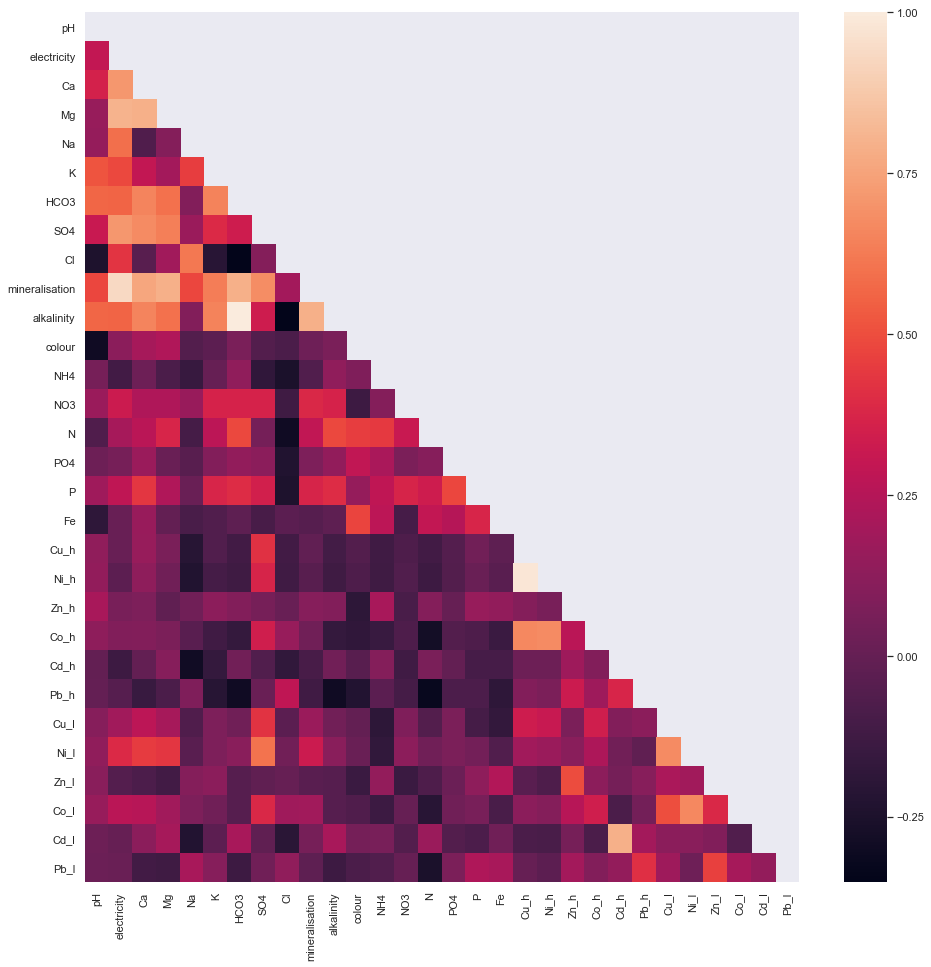
**Приложение Д**

Матрица корреляций для данных (с выбросами) по озёрам Арктического региона.

****

**Приложение Е**

Матрица корреляций для данных (без выбросов) по озёрам Арктического региона.

****

**Приложение Ж**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Озеро 1 | | | Озеро 2 | | Озеро 2 | | | Озеро 3 | | |
| MR1 | MR2 | MR3 | MR1 | MR2 | MR1 | MR2 | MR3 | MR1 | MR2 | MR3 |
| depth | -0.628 | -0.286 | -0.414 | -0.435 | -0.649 | -0.504 | -0.630 | -0.199 | -0.814 | 0.448 |  |
| Li | 1.010 |  |  | 0.871 | 0.182 | 0.861 | 0.182 | -0.203 | 1.003 | 0.156 |  |
| Be | 0.940 |  |  | 0.826 | 0.249 | 0.847 | 0.244 |  | 0.340 | 0.766 | 0.458 |
| P | -0.839 | 0.369 | -0.134 | -0.289 | -0.660 | -0.323 | -0.645 | 0.102 |  | -0.164 | 0.934 |
| Sc | 1.010 |  |  | 0.383 | 0.268 | 0.394 | 0.263 |  | -0.706 | 0.499 | -0.160 |
| Ti | 1.007 |  |  | 0.892 | 0.156 | 0.881 | 0.157 | -0.195 | 0.955 | 0.236 |  |
| V | 0.963 |  | 0.114 | 0.540 | 0.330 | 0.552 | 0.319 | -0.121 | 0.633 | -0.107 | 0.571 |
| Cr | 0.840 | 0.347 | -0.402 | 0.760 | 0.254 | 0.774 | 0.246 |  | 0.987 | 0.145 |  |
| Mn | 0.398 | 0.190 | 0.688 | -0.253 | -0.710 | -0.289 | -0.694 | 0.109 | 0.864 | -0.334 |  |
| Co | 0.828 |  | 0.296 | 0.732 | -0.179 | 0.798 | -0.200 | 0.527 | 0.751 |  | 0.518 |
| Ni | 0.947 |  | 0.130 | 0.506 | 0.439 | 0.528 | 0.423 | -0.109 | 0.849 |  | 0.316 |
| Cu | 0.356 | 0.124 | 0.688 | 0.141 | 0.726 | 0.205 | 0.695 |  | 0.339 | 0.425 | 0.729 |
| Zn | 0.148 | 0.771 |  | 0.148 | 0.496 | 0.210 | 0.473 | 0.208 | 0.646 | -0.288 | 0.478 |
| Rb | 1.010 |  |  | 0.924 | 0.101 | 0.908 | 0.104 | -0.199 | 1.010 |  |  |
| Sr | 0.913 |  | 0.213 | 1.020 | -0.287 | 1.014 | -0.278 | 0.114 | 0.884 | -0.102 | 0.261 |
| Y | 0.978 |  |  | 0.973 |  | 0.985 |  |  | -0.414 | 0.871 |  |
| Zr | 1.000 |  |  | 0.770 | 0.191 | 0.790 | 0.189 |  | 0.935 |  |  |
| Nb | 1.014 |  |  | 0.891 | 0.162 | 0.882 | 0.162 | -0.175 | 0.979 | 0.245 | -0.143 |
| Mo | 0.508 | 0.551 | 0.192 |  | 0.614 |  | 0.586 | 0.200 | 0.502 |  | 0.740 |
| Ag | -0.371 | 0.384 | 0.451 | -0.286 | -0.321 | -0.236 | -0.339 | 0.589 | -0.201 | 0.447 | 0.482 |
| Cd | -0.891 | 0.518 | 0.352 | -0.157 | 0.642 |  | 0.648 | 0.531 | 0.762 | -0.428 | 0.148 |
| Sn | 0.657 | 0.551 | -0.235 |  | 0.523 |  | 0.522 | 0.502 | 0.892 |  | 0.243 |
| Sb | -0.186 | 0.950 |  | -0.165 | 0.408 |  | 0.436 | 0.720 | 0.953 | -0.126 |  |
| Cs | 1.013 |  |  | 0.898 | 0.144 | 0.886 | 0.145 | -0.195 | 1.015 |  | -0.160 |
| Ba | 0.926 |  | 0.204 |  | -0.764 |  | -0.759 | 0.236 | 0.929 |  | 0.117 |
| La | 0.995 |  |  | 1.045 | -0.168 | 1.041 | -0.161 |  | 0.120 | 0.987 |  |
| Ce | 0.997 |  |  | 1.053 | -0.291 | 1.045 | -0.282 | 0.102 | 0.525 | 0.828 | -0.197 |
| Pr | 0.997 |  |  | 1.027 | -0.117 | 1.022 | -0.110 |  | 0.590 | 0.754 | -0.293 |
| Nd | 0.997 |  |  | 1.031 | -0.111 | 1.027 | -0.105 |  | 0.662 | 0.660 | -0.351 |
| Sm | 0.993 |  |  | 1.030 | -0.110 | 1.026 | -0.104 |  | -0.172 | 0.813 | -0.194 |
| Eu | 1.004 |  |  | 1.049 | -0.216 | 1.045 | -0.208 |  |  | 0.880 | -0.253 |
| Gd | 0.977 |  |  | 1.018 |  | 1.015 |  |  |  | 0.965 |  |
| Tb | 1.008 |  |  | 0.974 |  | 0.980 |  |  | -0.483 | 0.809 |  |
| Dy | 0.992 |  |  | 0.976 |  | 0.978 |  |  | -0.607 | 0.762 |  |
| Ho | 0.985 |  |  | 0.955 |  | 0.965 |  |  | -0.624 | 0.744 |  |
| Er | 0.986 |  |  | 0.960 |  | 0.960 |  |  | -0.536 | 0.796 |  |
| Tm | 0.962 |  |  | 0.923 | 0.101 | 0.925 | 0.101 |  | -0.534 | 0.759 |  |
| Yb | 0.981 |  |  | 0.947 |  | 0.948 |  |  | -0.134 | 0.837 | -0.223 |
| Lu | 0.990 |  |  | 0.899 | 0.160 | 0.903 | 0.158 |  | 0.283 | 0.896 | -0.170 |
| Hf | 1.011 |  |  | 0.843 | 0.205 | 0.852 | 0.203 |  | 0.941 |  |  |
| Ta | 1.014 |  |  | 0.905 |  | 0.895 |  | -0.139 | 0.986 | 0.106 |  |
| W | 0.679 | 0.548 | 0.183 | 0.159 | 0.639 | 0.213 | 0.618 |  | 0.838 |  | 0.388 |
| Tl | 0.996 |  | -0.102 | 0.767 | 0.293 | 0.786 | 0.286 |  | 0.964 |  | 0.107 |
| Pb | 0.342 | 0.719 | 0.233 | -0.196 |  | -0.114 |  | 0.783 | 0.965 | -0.154 |  |
| Bi | 0.764 | 0.385 | -0.101 |  |  |  |  | 0.698 | 0.971 | -0.139 |  |
| Th | 1.011 |  | -0.105 | 0.928 | 0.104 | 0.915 | 0.106 | -0.181 | 0.710 | 0.607 | -0.330 |
| U | 0.831 | 0.103 | 0.326 | 0.737 | 0.380 | 0.746 | 0.371 | -0.167 |  | 0.860 | 0.167 |